CS 6840: Algorithmic Game Theory **Spring 2017**

Lecture 11: February 17

Lecturer: Éva Tardos *Scribe: Austin Lin*

11.1 Games with learning

Consider the following game:

- Players 1...k
- Strategies s_i for player i
- Cost for player i: $c_i(\bar{s})$ where $\bar{s} = (s_1, ..., s_k)$

We play this game repeatedly for steps $1, 2, ..., T$. If i chooses s_i and everyone else chooses $\bar{s}_{-i}$, then the cost for i is $c_i(\bar{s})$. We want to consider a player that is learning from his/her experience. One question is how much information the player gets each step. We call this the **feedback** for the player.

Today we will be dealing with an easier version of feedback with increased information: $c_i(s_i', \bar{s}_{-i})$.

Definition 11.1 *Assuming the player learns $c_i(s_i', \bar{s}_{-i})$ for all her possible strategies s_i' is called **full-information feedback***

Online learning:

- One learner selects $s \in S$ (since we're only looking at one player, we drop the index i)
- Player gets feedback $c^t(s) = c_i(s, \bar{s}^t_{-i})$

11.2 Analyzing online learning games

Let's try selecting the best s at each step which gives us $\sum_t min_s(c^t(s))$. This standard is too hard to achieve: if the other players select at random, it can be the case that at each step one of $c(s_1)$ and $c(s_2)$ is 0 the other one is 1 at random. Now no matter what the player chooses, here expected cost is 1/2 at each step, while the optimum above is 0. But to reach this optimum, the player would need to know the outcome of random choices on the other players. This is hopeless.

A natural strategy here is to select s at time T that minimizes $\sum_{t=1}^{T-1} c(s)$.
This strategy does not work because it is deterministic and, in the worst case, the other players could play to maximize my cost given my choice of strategy.

11.2.1 Approximate no regret learning

Looking at our Nash or coarse correlated equilibrium definition may give us an idea of a standard that is both doable and useful. Recall the Nash and CCE conditions:

If we have a probability distribution σ, then the Nash condition gives us the inequality:

$$\mathbb{E}_{s\sim\sigma}[c_i(s)] \leq \mathbb{E}_{s\sim\sigma}[c_i(s_i', \bar{s}_{-i}]\forall i, s_i'$$

So while players at Nash or a coarse correlated Nash mix between many strategies, we are comparing them to one fixed strategy. The analog for learning is the following: Let's try selecting a single strategy to do almost as well as $min_s(\sum_t c^t(s))$.

Approx no regret learning:

- With strategies over time $s^1, ..., s^T$, we get cost $\sum_t c^t(s^t)$
- OPT in this sense for the player is $min_s \sum_t c^t(s^t)$, this is the best single strategy (with hindsight).
- If we define ϵ as the **learning error**, then we say:
  Approx regret $= \sum_t c^t(s^t) - (1+\epsilon)\min_s \sum_t c^t(s^t)$.
  Next time we will show that there are simple (and natural) algorithms) that achieve approx regret $\leq O(\frac{log|S|}{\epsilon})$.

Suppose all players do approx no regret learning (using same ϵ), and assume the game is (λ, μ) smooth.
In each step t, the strategies are represented by $\bar{s}^t = (s_1^t, ..., s_k^t)$, and across all steps we have vectors $(\bar{s}^1, ..., \bar{s}^T)$.
For a player i that does learning, we get the inequality:

$$\sum_t c_i(\bar{s}^t) \leq (1+\epsilon) min_{s_i'} \sum_t c_i(s_i', \bar{s}_{-i}^t) + O(\frac{log|S|}{\epsilon})$$

First, we can connect the outcome of such a learning process by players to the notion of coarse correlated equilibrium we have seen already. To do this, we think of $\bar{s}^1, ..., \bar{s}^T$ as a probability distribution (pick a random time t, and evaluate the strategy vector $(s_1^t, \ldots, s_k^t)$, this defined the empirical probability distribution $\bar{\sigma}$.
Now we can combine our two inequalities to get:
$\mathbb{E}_{\bar{s}\sim\bar{\sigma}}[c_i(\bar{s})] = \frac{1}{T}\sum_t c_i(\bar{s}^t)$
$\mathbb{E}_{\bar{s}\sim\bar{\sigma}}[c_i(s_i', \bar{s}_{-i})] = \frac{1}{T}\sum_t c_i(s_i', \bar{s}_{-i}^t)$
Substituting into our inequality gives us:

$$\mathbb{E}_{\bar{s}\sim\bar{\sigma}}[c_i(\bar{s})] \leq (1+\epsilon)\mathbb{E}_{\bar{s}\sim\bar{\sigma}}[c_i(s_i', \bar{s}_{-i})] + O(\frac{log|S|}{\epsilon T})$$

So learning outcomes are very close to a coarse correlated equilibria, in the sense of the inequality required for such equilibria is approximately true (with a small additive and a small multiplicative error).

11.2.2 Price of Anarchy bounds

Given that the empirical distribution of the learning outcome is so close to coarse correlated equilibria, we expect that they also satisfy the price of anarchy consequence with a small added error. Here is what we get:

Theorem 11.2 *If a game is (λ, μ) smooth, then players are all using approximately no-regret learning with the same error parameter ϵ than the average cost*

$$\sum_t c(\bar{s}^t) \leq \frac{\lambda(1+\epsilon)}{1-\mu(1+\epsilon)} T \min_{\bar{s}^*} \sum_i c_i(\bar{s}^*) + O(\frac{k(log|S|)}{\epsilon(1-\mu(1+\epsilon))})$$

Proof: Using that all players do approximately no-regret learning, and using $s'_i = s^*_i$, the ith coordinate of the strategy vector s^* that minimizes $\min_{\bar{s}^*} \sum_i c_i(\bar{s}^*)$, we get

$$\textstyle\sum_t c_i(\bar{s}^t) \leq (1+\epsilon) \sum_t c_i(s^*_i, \bar{s}^t_{-i}) + O(\frac{log|S|}{\epsilon}))$$

Adding these up for all players and using smoothness we get:

$$\begin{aligned} \sum_t \sum_i c_i(s^t) &\leq (1+\epsilon) \sum_t \sum_i c_i(s^*_i, \bar{s}^t_{-i}) + O(k\frac{log|S|}{\epsilon})) \\ &\leq (1+\epsilon) \sum_t (\lambda \sum_i c_i(\bar{s}^*) + \mu \sum_i c_i(\bar{s}^t)) + O(\frac{k(log|S|)}{\epsilon}) \end{aligned}$$

Rearranging gives us:

$$(1-(1+\epsilon)\mu) \sum_t c_i(s^t) \leq \lambda(1+\epsilon) T * OPT + O(k\frac{(log|S|)}{\epsilon})$$

as $\sum_i c_i(\bar{s}^*) = OPT$ occurs at each step t.
Now dividing both sides by $(1-(1+\epsilon)\mu)$ gives us the bound claimed. ■